Article:
Bonvoisina, Toby, Paton, Lewis William orcid.org/0000-0002-3328-5634, Hewitt, Catherine Elizabeth orcid.org/0000-0002-0415-3536 et al. (3 more authors) (2020) Collaborative care for depression in older adults:how much is enough? Behaviour research and therapy. 103725. ISSN: 1873-622X

https://doi.org/10.1016/j.brat.2020.103725

Collaborative care for depression in older adults: how much is enough?

Toby Bonvoisin[a, b], Lewis W Paton[a], Catherine Hewitt[a], Dean McMillan[a, c], Simon Gilbody[a, c], Paul A Tiffin[a, c]

[a]: Department of Health Sciences, University of York, Heslington, YO10 5DD, UK

[b]: Hull University Teaching Hospitals NHS Trust, Hull, HU3 2JZ, UK

[c]: Hull York Medical School, University of York, Heslington, YO10 5DD, UK

Email addresses: ttb508@york.ac.uk, lewis.paton@york.ac.uk, catherine.hewitt@york.ac.uk, dean.mcmillan@york.ac.uk, simon.gilbody@york.ac.uk, paul.tiffin@york.ac.uk

Correspondence to: lewis.paton@york.ac.uk, Department of Health Sciences, University of York, Heslington, YO10 5DD, UK

ABSTRACT

Collaborative care in primary care has been shown to be effective for subthreshold depression in older adults in the 'CASPER' trial. However, to understand the impact of adherence, and to explore the minimum effective dose of collaborative care, we reanalysed the trial data using a complier average causal effect (CACE) analysis. Data were available for 705 participants, 519 with 12-month PHQ-9 scores. 'Compliance' could be observed for participants in the intervention group. Latent complier status in the control group was estimated. Completion of five or more sessions of care was defined as 'compliance'. Sensitivity analyses, using alternative cut-offs of two to eight sessions, assessed the impact of changing the definition of 'compliance'. Compliers in the intervention group had lower PHQ-9 scores at 12-month follow up than assumed compliers in the control group (1.75 lower, 95% confidence interval 0.29 to 3.21, p=0.02), a greater effect than originally reported. Sensitivity analyses confirmed statistically significant differences between the intervention and control groups in those attending five or more sessions. We conclude that collaborative care is causally effective in reducing subthreshold depressive symptoms in older people who adhere to treatment. Our findings suggest the minimum effective dose is five sessions.

Keywords late life depression; primary health care; treatment; intervention; causal analysis

INTRODUCTION

Subthreshold depression, defined as the presence of depressive symptoms not sufficient enough to qualify for a diagnosis of depressive disorder (Rodríguez, Nuevo, Chatterji, & Ayuso-Mateos, 2012), is common in older adults. It impairs quality of life (Chachamovich, Fleck, Laidlaw, & Power, 2008) and is a risk factor for progression to more severe ('case level') depressive disorder (Cuijpers & Smit, 2004). There is little evidence to guide management (Cameron, Reid, & MacGillivray, 2014; Cuijpers, Smit, & Van Straten, 2007). One credible approach is 'collaborative care', which uses case management and low intensity interventions, such as behavioural activation (Lewis et al., 2017). The CASPER trial (CollAborative care and active surveillance for Screen-Positive EldeRs with subthreshold depression) evaluated this approach against usual primary care (Gilbody et al., 2017; Lewis et al., 2017).

The CASPER trial consisted of 705 patients aged 65 years or older from the North of England identified as having subthreshold depression via the Mini International Neuropsychiatric Interview. Participants were excluded from the trial if identified by their GP as having known alcohol dependency, psychotic symptoms, the presence of comorbidities (such as suicidal risk) or had experienced other recent life events (such as bereavement). Eligible patients were randomised equally to one of two trial arms. Those in the control group (n=361) received usual primary care for management of subthreshold depression. Those randomised to the intervention group (n=344) received targeted collaborative care, consisting of support and symptom tracking (Pasterfield et al., 2014), as well as being offered a course of behavioural activation (Ekers et al., 2014). This was delivered by a case manager, and planned to be weekly over a period eight weeks. The first delivered session was planned to be face-to-face, with subsequent sessions delivered via telephone. However, the ratio of face-to-face and telephone sessions varied depending on the circumstances of the patient. Of the planned eight sessions, the median number of total sessions received was seven, with a range of one to fifteen. Sixty-seven (19.5%) participants in the intervention arm of the original trial received more than eight sessions of care based on agreed clinical need.

Outcome data were collected at baseline, and 4- and 12-months post randomisation. Alongside baseline demographic data a number of health outcomes were collected at each time point. The primary outcome was the participant's depressive symptoms, as evaluated by

the Patient Health Questionnaire (PHQ-9) score. At 12 months post randomisation, PHQ-9 data were available for 519 participants (n=235 in the treatment arm, n=284 in the control arm). Using intention-to-treat (ITT) analysis, those randomised to collaborative care had lower scores compared to the usual care group (-1.33 points, 95% confidence interval (CI) -2.10 to -0.55). The relative risk of participants in the intervention group exceeding the screening threshold for depression was 0.65 (95% CI 0.46 to 0.91) compared to the control group. However, this ITT analysis does not account for the effects of adherence to the intervention on the outcome, potentially underestimating the causal effects of treatment for those who receive an adequate 'dose' of collaborative care. Moreover, it is unknown what the minimum effective 'dose' is, in terms of session numbers, in order to gain a meaningful advantage over care as usual. This information would be useful in clinical practice.

Those who adhere to treatment may systematically differ from those who do not. Grouping patients by actual treatment received rather than intended treatment ('as-treated' analysis) or dropping all participants who did not follow their allocated protocol ('per-protocol' analysis) potentially introduces bias, removing the benefits of randomisation in controlling for unmeasured factors (McNamee, 2009). Estimation of the 'complier average causal effect' (CACE) is a method whereby the benefits of randomisation are preserved and provides an unbiased estimate of the effect of the intervention for those who comply with the intervention (Peugh, Strotman, McGrady, Rausch, & Kashikar-Zuck, 2017). In our methodology, we use the term 'compliance' rather than 'adherence' for consistency with the CACE literature. We therefore undertook a secondary analysis of the CASPER data using this approach. We also evaluated the minimum number of sessions (these included telephone and face-to-face contacts) required to produce a significant positive difference in reported symptoms, compared to care as usual, unlikely to be due to chance alone.

METHODS

We performed a secondary analysis of data from the 705 participants in the original trial (Lewis et al., 2017). The participants' 12-month PHQ-9 score was the primary outcome. Our secondary outcome was binary; whether the 12-month PHQ-9 score exceeded the threshold of ≥10 points, consistent with moderate to severe depression. We assumed participants were of two compliance types: "non-compliers" (those who would never comply with treatment even if referred) and "compliers" (those who would adhere to treatment if they had access to

it). This variable is observable in the intervention group, but not in the control. We used a mixture model to estimate latent complier status from baseline variables in those randomised to the control group. We then performed a CACE analysis to compare outcomes for compliers in the intervention group with outcomes of those in the control group who we predict would have complied with the treatment if given the opportunity to do so. Variables included in the mixture model were gender, age at randomisation, baseline PHQ-9 score, baseline Generalised Anxiety Disorder seven-item scale (GAD-7), baseline Patient Health Questionnaire-15 (PHQ-15) somatic symptom severity scale score, baseline responses to the Whooley questions, presence of comorbidities, if the patient had continued education after minimum school leaving age, and if the participant had a degree or equivalent professional qualification. Analyses were conducted using Mplus version 8.1 (www.statmodel.com).

For participants in the intervention arm, the number of sessions of collaborative care they received could be observed. However, it was not possible to distinguish between face-to-face sessions and those conducted by telephone, and thus it had to be assumed that these sessions were equivalent in their effectiveness. Participants who completed five or more sessions were defined as compliers based on an *a priori* clinically-informed judgement following discussions with clinicians involved in the original CASPER trial regarding what the 'minimum effective dose' would be in terms of number of sessions. Sensitivity analyses were conducted using alternative cut-offs of two to eight sessions to assess the impact of changing the definition of 'compliance' on the relative effect size, compared to care as usual.

A further sensitivity analysis was performed using multiple imputation to generate predictor variable values and 12-month outcomes for participants where these were missing. Chained equations, as implemented in Stata version 14.2, were used. This approach uses a set of conditional models to impute data in a series of variables where there are missing values. Multiple datasets are created to allow for uncertainty in the imputed values. Variables were included in the imputation model if they displayed an association with PHQ-9 scores on univariable regression analysis. Missing data were imputed from baseline demographic variables, and baseline and four-month outcomes for PHQ-9, European Quality of Life-5 Dimensions (EQ-5D), GAD-7, quality of life as measured by the Short Form questionnaire-12 items (SF-12), the Whooley questions, and the presence of some physical comorbidities ('diabetes', 'heart disease' and 'other'). Thirty imputed datasets were created in Stata, which were then imported into Mplus before CACE estimates were calculated and pooled. The results on non-imputed and imputed datasets were then compared.

The validity of the CACE estimation rests on a number of assumptions (Dunn, 2002; Frangakis & Rubin, 1999): treatment was randomised, outcomes for individual participants were statistically independent of each other (the 'stable unit treatment value assumption' (SUTVA)), no participants in the control group accessed the treatment, and treatment allocation had no influence on non-compliers (the 'exclusion restriction').

RESULTS

Data were available for the 705 participants of the CASPER trial. Of these, 686 had complete observations on all baseline and demographic variables included in the mixture model. Missing data in baseline variables were observed only for baseline PHQ-9 score (n=7), baseline GAD-7 score (n=7), baseline PHQ-15 score (n=10), continuation of education post-school (n=5) and if the patient had a degree or equivalent qualification (n=8). Twelve-month PHQ-9 outcomes were available for 519 of the 705 participants across both arms of the trial (73.6%), and specifically 235/344 (68.3%) in the treatment arm and 284/361 (78.7%) in the control arm. Complier status could be observed in the 344 cases randomised to treatment in the original CASPER trial, of which 202 attended five or more sessions (58.7%). Complier status (latent or observed) could be estimated in those with complete data (i.e. 686 of the 705 cases).

When defining 'compliance' as completing five or more sessions, compliers in the treatment group had statistically significantly lower PHQ-9 scores than assumed compliers in the control group, by 1.75 points (95% CI 0.29 to 3.21, p=0.019). Those who complied with collaborative care had relatively lower odds of reporting a PHQ-9 score above the moderate/severe depression threshold at 12 months than assumed compliers in the control arm (OR 0.18, 95% CI 0.04 to 0.76, p=0.019). As seen in figure 1, the results of our sensitivity analyses confirmed that five sessions of collaborative care is the 'minimum effective dose' likely to result in statistically significant improvement, compared to care as usual.

In terms of the results from the imputed data; an average decrease of 1.38 points on the PHQ-9 (95% CI 0.03 to 2.72, p=0.045) was observed between compliers and assumed compliers. For the binary outcome, no statistically significant results were observed. We note a weak trend that compliers in the treatment group had lower odds of being above the screening threshold for depression than assumed compliers in the control group (OR 0.93,

95% CI 0.39 to 2.95, p=0.32). However, this was not statistically significant at the p=0.05 level.

DISCUSSION

Our results suggest that collaborative care is causally effective in reducing subthreshold depressive symptoms in older people adhering to treatment. Unsurprisingly, our causal estimate for the effect of intervention on symptoms for 'compliers' is modestly larger than that from the original ITT analysis. Importantly, the findings from our sensitivity analyses suggest that five or more sessions was the minimum effective dose, in order to achieve effects statistically significantly superior to care as usual. Moreover, the difference in outcome attributable to the intervention did not appear to increase when the definition of compliance was increased beyond five sessions.

Our CACE estimate provides a reasonable appraisal of the likely impact of collaborative care on patients who adhere. Data were relatively complete and the results from the imputed datasets for our primary outcome (continuous PHQ-9 score at 12 months post-randomisation) were in the same direction as those from the non-imputed data, as well as remaining statistically significant. Therefore, we can be reasonably confident in relation to our estimate of the causal impact of collaborative care on continuous PHQ-9 scores at 12-month follow up. However, the effect-size estimated from the imputed data, in relation to whether participants were above the screening (binary) threshold for PHQ-9 at follow up did not reach statistical significance, in contrast to the results from the non-imputed data. Several imputation models were explored, and returned similar results. It is worth reflecting on the possible explanations for this disparity. The most plausible reason is partly due to the inclusion criteria of the original trial. Participants eligible for the trial were identified as having subthreshold depression and not major depressive disorder on structured diagnostic interview (the Mini International Neuropsychiatric Interview). When depressive symptoms were subsequently measured by PHQ-9 scores, many participants therefore had PHQ-9 scores close to the threshold cut-off of ten points for subthreshold vs major depression. As a result, relatively small changes in PHQ-9 scores can change the binary classification of subthreshold depression or clinically depressed. This uncertainty, over which side of the threshold an individual would have been at follow up, may have amplified uncertainty introduced by the imputation, explaining the subsequent disparities in the CACE estimates

relating to this outcome between imputed and non-imputed data. As such, our CACE estimate of the impact of treatment on binary depression status is less certain, and results for our secondary outcome should be treated with some caution.

The assumptions underpinning the CACE analysis are likely to have been generally met. Participants were randomly allocated and those in the control group would not have accessed collaborative care. The SUTVA assumption may not hold, given that participants may have shared case managers. However, the differences across case managers were reported as relatively trivial, with an intracluster correlation within case managers of less than 0.01 (Lewis et al., 2017). Violation of the exclusion restriction can lead to invalid CACE estimates. However, it has previously been shown that inclusion of baseline covariates can reduce bias in the CACE estimation from violation of the exclusion restriction, if an association between those covariates and compliance exists (Jo, 2002). We included a number of baseline variables in our modelling process, and it is implausible that there is zero association between these variables and compliance. As such, we are confident that our findings are reasonably robust to any potential departures from the exclusion criteria. Other limitations include using PHQ-9 as an outcome measure, which was originally designed as a screening measure (Cameron et al., 2011; Kroenke & Spitzer, 2002), as well as the assumption of equivalence in effectiveness between face-to-face and telephone sessions of collaborative care.

Nevertheless, the findings of this study are valuable to providers and commissioners of mental health services for older people. They indicate the likely symptomatic benefits to those who adhere to care and the minimum required number of sessions (namely, five) needing to be provided in order to realise this in most cases. In this trial, 202 participants out of 344 (58.7%) attended the minimally effective dose of collaborative care. The Improving Access to Psychological Therapies (IAPT) programme, one of the principal avenues for accessing talking therapies for depression in England, had a completion rate for treatment of 38.9% in 2018/19 (NHS Digital, 2019). While not directly comparable treatments or statistics, this suggests that engagement in collaborative care in this trial was reasonable. Nevertheless, improving adherence rates for collaborative care further has clear benefits. The CASPER trial reported a number of potential reasons for non-engagement with the intervention (Lewis et al., 2017). These included a perceived lack of need for the intervention from patients, difficulty with the intervention itself, and raised awareness of symptoms and thus deciding

that the intervention either wasn't necessary or that it was actually worsening symptoms. Pragmatic difficulties with delivery of collaborative care, such as unavailable case managers, was also quoted as a reason for lack of engagement. Our results suggest the relative benefits of receiving a dosage of five sessions of care, and thus further research should investigate strategies for improving adherence in those offered collaborative care and seek to identify the most effective components of this approach.

Figure legends

Figure 1. Complier average causal effect (CACE) of collaborative care on 12-month outcomes for varying numbers of minimum sessions (telephone and face-to-face) to define 'compliance'.

Ethics statement

This study used anonymised data from the CASPER trial that was previously approved by the NHS Leeds East Ethics Committee (10/H1306/61). Additional ethics approval was not required.

Consent statement

All participants were aged 65 years or older and gave written informed consent to participate in this CASPER trial between March 2011 and July 2013.

Funding

The data used in this study was collected as part of the CASPER trial, which was funded by the UK National Institute of Health Research Health Technology Assessment Programme (project number 08/19/04).

TB was funded by the NHS Academic Foundation Programme, Yorkshire and Humber deanery. PAT is funded via an National Institute for Health Research (NIHR) Career Development Fellowship. This report presents independent research funded by the NIHR. The views expressed are those of the author(s) and not necessarily those of the (partner organisation), the NHS, the NIHR or the Department of Health and Social Care.

Data Availability

Data used in this study are available from the University of York Trials Unit upon application. Code used in the analysis is available from the corresponding author.

Acknowledgements

None

References

Cameron, I. M., Cardy, A., Crawford, J. R., du Toit, S. W., Hay, S., Lawton, K., . . . Winning, S. (2011). Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II. *British Journal of General Practice, 61*(588), e419-e426.

Cameron, I. M., Reid, I. C., & MacGillivray, S. A. (2014). Efficacy and tolerability of antidepressants for sub-threshold depression and for mild major depressive disorder. *Journal of Affective Disorders, 166*, 48-58.

Chachamovich, E., Fleck, M., Laidlaw, K., & Power, M. (2008). Impact of major depression and subsyndromal symptoms on quality of life and attitudes toward aging in an international sample of older adults. *The Gerontologist, 48*(5), 593-602.

Cuijpers, P., & Smit, F. (2004). Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. *Acta Psychiatrica Scandinavica, 109*(5), 325-331.

Cuijpers, P., Smit, F., & Van Straten, A. (2007). Psychological treatments of subthreshold depression: a meta‐analytic review. *Acta Psychiatr Scand, 115*(6), 434-441.

Dunn, G. (2002). Estimating the causal effects of treatment. *Epidemiology Psychiatric Sciences, 11*(3), 206-215.

Ekers, D., Webster, L., Van Straten, A., Cuijpers, P., Richards, D., & Gilbody, S. (2014). Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. *PLoS One, 9*(6).

Frangakis, C. E., & Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika, 86*(2), 365-379.

Gilbody, S., Lewis, H., Adamson, J., Atherton, K., Bailey, D., Birtwistle, J., . . . Ekers, D. (2017). Effect of collaborative care vs usual care on depressive symptoms in older adults with subthreshold depression: the CASPER randomized clinical trial. *JAMA, 317*(7), 728-737.

Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of interventions with non ‐ compliance. *Statistics in Medicine, 21*(21), 3161-3181.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509-515.

Lewis, H., Adamson, J., Atherton, K., Bailey, D., Birtwistle, J., Bosanquet, K., . . . Foster, D. (2017). CollAborative care and active surveillance for Screen-Positive ElderRs with subthreshold depression (CASPER): a multicentred randomised controlled trial of clinical effectiveness and cost-effectiveness. *Health Technology Assessment Reports, 21*(8), 1.

McNamee, R. (2009). Intention to treat, per protocol, as treated and instrumental variable estimators given non ‐ compliance and effect heterogeneity. *Statistics in Medicine, 28*(21), 2639-2652.

NHS Digital. (2019). Annual report on the use of IAPT services, England 2018-19. Retrieved from https://files.digital.nhs.uk/1C/538E29/psych-ther-2018-19-ann-rep.pdf

Pasterfield, M., Bailey, D., Hems, D., McMillan, D., Richards, D., & Gilbody, S. (2014). Adapting manualized behavioural activation treatment for older adults with depression. *The Cognitive Behaviour Therapist, 7*.

Peugh, J. L., Strotman, D., McGrady, M., Rausch, J., & Kashikar-Zuck, S. (2017). Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer. *Journal of School Psychology, 60*, 7-24.

Rodríguez, M. R., Nuevo, R., Chatterji, S., & Ayuso-Mateos, J. L. (2012). Definitions and factors associated with subthreshold depressive conditions: a systematic review. *BMC Psychiatry, 12*(1), 181.